

Comparación BD NoSQL

Yaneth Virginia Aquino Huallpa, Arlyn Cotrado Coaquira, Sharon Sosa Bedoya, and Marlon Villegas Arando
Universidad Privada de Tacna \Facultad de Ingenieria \Escuela Profesional de Ingenieria de Sistemas

Resumen

En el presente artículo se relata la comparación hecha de bases de datos NoSQL, describiéndolas y analizando su importancia, así como las definiciones de tipos de bases de datos NoSQL, con el fin de proporcionar un punto de partida para los trabajos en esta área. Y también la creación de una base de datos, inserción y consultas de datos NoSQL mediante Docker.

Palabras clave: NoSQL, Bases de datos, Docker.

Abstract

In the present article the comparison made of NoSQL databases is described, describing them and analyzing their importance, as well as the definitions of NoSQL database types, in order to provide a starting point for work in this area. And also the creation of a database, insertion and queries of NoSQL data through Docker.

Keywords: NoSQL, Databases, Docker.

I. INTRODUCCIÓN

Día a día el manejo de la información se hace más complejo; diferentes factores hacen que las personas involucradas en el área busquen tecnologías que le ayuden con este problema. Las bases de datos relacionales son las mas comunes, pero en los últimos años ha aumentado el interés por las bases de datos NoSQL (Not only SQL), un nuevo conjunto de tecnologías que pueden contribuir al manejo de información.

Por lo anterior, el presente documento hace una revisión de las tecnologías NoSQL, haciendo posible hacer una comparación.

El resto de este artículo está organizado de la siguiente manera. En la Sección 2 se muestra los materiales y métodos usados para el desarrollo de este artículo. La Sección 3 se explican los resultados. Y finalmente, las conclusiones están en la Sección 4.

II. MATERIALES Y MÉTODOS

A. Materiales

- Virtualización activada en el BIOS
- Docker Desktop
- Windows 10 64bit: Pro, Enterprise o Education, con al menos 4GB de RAM.

B. Métodos

- Se utilizó como material artículos y libros relacionados a la base de datos NoSQL y sus tipos, así como páginas web.

III. MARCO TEÓRICO

A. Base de datos

- Una base de datos es una colección de datos organizados según un determinado criterio
- Estos datos se pueden leer, crear, actualizar y borrar
- También existen motores de base de datos que nos permiten hacer todas estas operaciones de forma más fácil..[6]

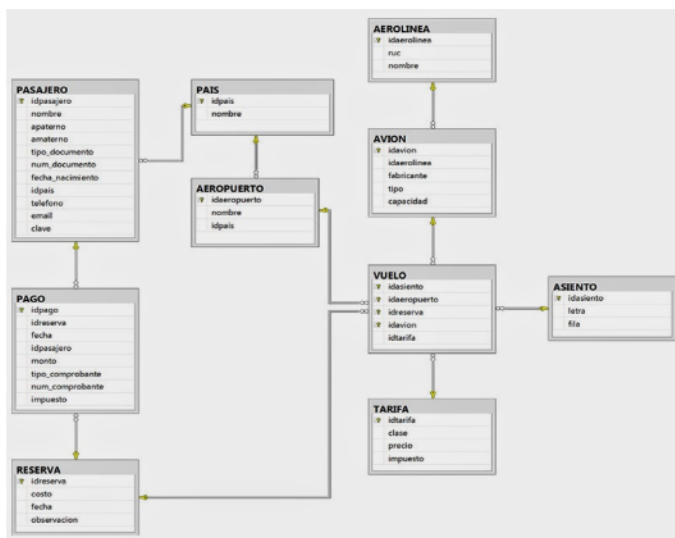
B. Tipos de base de datos

- Existen distintos tipos de bases de datos que se utilizan para solucionar distintos tipos de problemas
- Dentro de la gran familias de bases de datos podemos encontrar las del tipo base de datos relacionales y las no relacionales
- Las bases de datos relacionales se conocen generalmente como las SQL
- Las no relacionales se conocen como NoSQL

- Cada tipo de base de datos tiene beneficios y contras a la hora de almacenar, leer, actualizar o borrar los datos

C. Bases de datos relacional

- Desde su definición vemos que una base de datos puede ser relacional si cumple con algo conocido como el modelo relacional.
- En la definición de modelo relacional nos podemos quedar con la idea de tablas que tienen columnas para describir los datos que están relacionados entre si.[3]
- Ejemplo modelo relacional:[6]



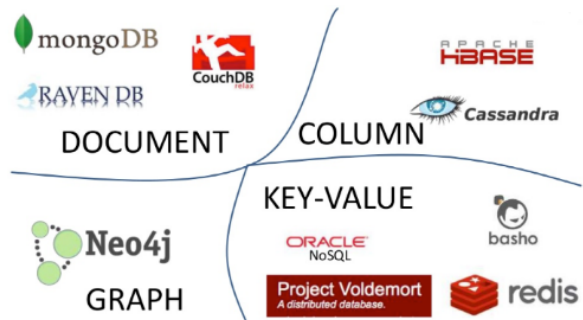
D. Modelo NoSQL

Este tipo de Base de datos es la respuesta a la necesidad de gestionar volúmenes masivos de información, de tal manera que surge la base de datos NoSQL, término que fue acuñado a finales de los años 90 y que engloba todas las tecnologías de almacenamiento estructurado que no cumplen el esquema relacional, mayormente usado.

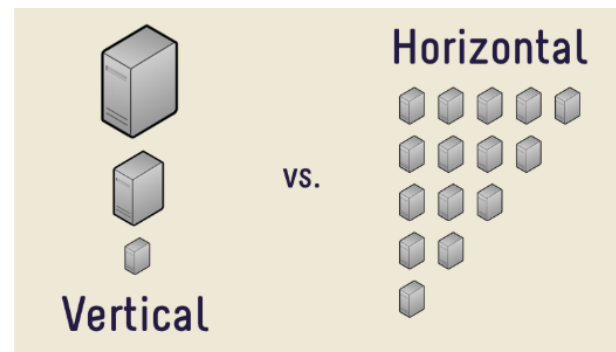
La cantidad de información manejada por las comunidades, redes sociales, buscadores, y muchos otros proyectos en el ámbito de la Web es enorme y abrumadora, lo que ha hecho que surjan nuevas arquitecturas de almacenamiento de información, que deben ser de alto rendimiento, escalables y distribuidas.

- Se conoce como NoSQL (Not Only SQL) al grupo de bases de datos que no son relacionales

- Dentro de esta clasificación se encuentran las bases de clave/valor, orientadas a documentos, grafos, de grandes columnas .[6]



- Este tipo de bases de datos escala de forma horizontal
- Podemos utilizar muchas máquinas chiquitas para crecer y satisfacer las necesidades de los negocios actuales.[6]



IV. RESULTADOS

A. Creacion de base de datos NoSQL con MongoDB

- MongoDB es una base NoSQL orientada a documentos
- Permite guardar documentos en formato de JSON
- Tiene esquema flexible, es decir que podemos cambiar la estructura de nuestros documentos sin ningún problema
- MongoDB está preparado para escalar fácilmente de manera horizontal
- Dado que aprendimos ECMAScript vamos a utilizar un motor de base de datos que nos permite seguir utilizando este lenguaje para guardar nuestros datos.

B. Instalar MongoDB en Docker

- Ingresar sus credenciales creadas en Docker Hub para iniciar sesión en el aplicativo. Ubicar la aplicación PowerShell, ejecutarla como Administrador. En la ventana de comandos de PowerShell escribir lo siguiente.



- Para instalar MongoDB primero tenemos que ejecutar el siguiente código.

```
PS C:\> docker pull mongo
Using default tag: latest
latest: Pulling from library/mongo
85b42117c431: Pull complete
ad9c569a8d98: Pull complete
293b44f45162: Pull complete
8c175077525d: Pull complete
4e73525b52ba: Pull complete
a22695a3f5e9: Pull complete
420eb4b7be5d: Pull complete
017ec49b70bf: Pull complete
26470656e2db: Pull complete
5fec59c7382a: Pull complete
12f57ad1345c: Pull complete
7102859e924d: Pull complete
555c1275dd6f: Pull complete
Digest: sha256:01dc9fb07aae875678047e2d8550beb6fc34b7e76c60a1e7d048f6700dead0
Status: Downloaded newer image for mongo:latest
```

- Verificar que el contenedor se este ejecutando correctamente mediante el comando:

```
PS C:\> docker ps
CONTAINER ID   IMAGE      COMMAND                  CREATED        STATUS        PORTS
mongo         mongo:3.5 "docker-entrypoint.s..." 35 minutes ago Up 35 minutes 27017/tcp
```

- Proceder a verificar la imagen con el siguiente comando:

```
PS C:\> docker images
REPOSITORY    TAG       IMAGE ID       CREATED        SIZE
mongo         latest   a3639b2a0f13   4 days ago    412MB
```

- Seguidamente ejecutar el comando. Como respuesta se visualizará un ID que corresponde al contenedor:

```
PS C:\> docker run --name virginia-mongo -d mongo
093a88bc0df784cd59774211cd23a78ecb5d83997b4fd260cd300d36543dd3
PS C:\>
```

- Para conectar a nuestro localhost

```
PS C:\> cd 'C:\Program Files\MongoDB\Server\'
```

- Para exponer ese puerto para nuestro que podamos acceder al contenedor

```
PS C:\Program Files\MongoDB\Server> ls

Directorio: C:\Program Files\MongoDB\Server

Mode                LastWriteTime         Length Name
----                -
d-----          23/06/2019     11:34             4.0
```

- MongoDB está preparado para escalar fácilmente de manera horizontal

```
PS C:\Program Files\MongoDB\Server\4.0> cd bin
PS C:\Program Files\MongoDB\Server\4.0\bin> ls

Directorio: C:\Program Files\MongoDB\Server\4.0\bin

Mode                LastWriteTime         Length Name
----                -
-a-----          28/05/2019     21:36    10249411 bsondump.exe
-a-----          28/05/2019     21:59      1568 InstallCompass.ps1
-a-----          3/04/2018      10:58    2462720 libeay32.dll
-a-----          28/05/2019     21:59    18461184 mongo.exe
-a-----          28/05/2019     21:34       570 mongod.cfg
-a-----          28/05/2019     22:01    32702976 mongod.exe
-a-----          28/05/2019     22:01    360729776 mongod.pdb
-a-----          28/05/2019     21:41    19350603 mongodump.exe
-a-----          28/05/2019     21:38    18841845 mongoexport.exe
-a-----          28/05/2019     21:38    18683050 mongofiles.exe
-a-----          28/05/2019     21:39    19029681 mongoimport.exe
-a-----          28/05/2019     21:40    18419576 mongorestore.exe
-a-----          28/05/2019     21:58    16987264 mongos.exe
-a-----          28/05/2019     21:58    188887040 mongos.pdb
-a-----          28/05/2019     21:37    18930780 mongostat.exe
-a-----          28/05/2019     21:41    18518542 mongotop.exe
-a-----          3/04/2018      10:58    3578888 sslseay32.dll
```

- MongoDB está preparado para escalar fácilmente de manera horizontal

```
PS C:\Program Files\MongoDB\Server\4.0\bin> docker info virginia-mongo
'docker info' accepts no arguments.
See 'docker info --help'.

Usage: docker info [OPTIONS]

Display system-wide information
```

- MongoDB está preparado para escalar fácilmente de manera horizontal

```
PS C:\Program Files\MongoDB\Server\4.0\bin> docker inspect virginia-mongo
[
  {
    "Id": "093a88bc0df784cd59774211cd23a78ecb5d83997b4fd260cd300d36543dd3",
    "Created": "2019-06-23T16:44:02.2296357Z",
    "Path": "docker-entrypoint.sh",
    "Args": [
      "mongo"
    ],
    "State": {
      "Status": "running",
      "Running": true,
      "Paused": false,
      "Restarting": false,
      "OOMKilled": false,
      "Dead": false,
      "Pid": 2547,
      "ExitCode": 0,
      "Error": "",
      "StartedAt": "2019-06-23T16:44:06.209002Z",
      "FinishedAt": "0001-01-01T00:00:00Z"
    },
    "Image": "sha256:a3639b2a0f136e0400a2639c6150b7f57a0947597714277c956762b5a0a042",
    "ResolvConfPath": "/var/lib/docker/containers/093a88bc0df784cd59774211cd23a78ecb5d83997b4fd260cd300d36543dd3/resolv.conf",
    "HostnamePath": "/var/lib/docker/containers/093a88bc0df784cd59774211cd23a78ecb5d83997b4fd260cd300d36543dd3/hostname",
    "HostsPath": "/var/lib/docker/containers/093a88bc0df784cd59774211cd23a78ecb5d83997b4fd260cd300d36543dd3/hosts",
    "LogPath": "/var/lib/docker/containers/093a88bc0df784cd59774211cd23a78ecb5d83997b4fd260cd300d36543dd3/093a88bc0df784cd59774211cd23a78ecb5d83997b4fd260cd300d36543dd3-json.log"
  }
]
```

- En PowerShell ejecutar el siguiente comando ,Verificar la eliminación del contenedor con ejecutando

```
PS C:\Program Files\MongoDB\Server\4.0\bin> docker stop virginia-mongo
virginia-mongo
PS C:\Program Files\MongoDB\Server\4.0\bin> docker rm virginia-mongo
virginia-mongo
```

- Seguidamente ejecutar el comando: Como respuesta se visualizará un ID que corresponde al contenedor

```
PS C:\Program Files\MongoDB\Server\4.0\bin> docker run -p 27017:27017 --name virginia-mongo -d mongo
11e5f2ab785172f6c8d6b01d127c6c1dc672ab40764f5a9b2
```

- Para conectar a nuestro localhost

```
PS C:\Program Files\MongoDB\Server\4.0\bin> .\mongo.exe
MongoDB Shell version 4.0.10
connecting to: mongodb://127.0.0.1:27017/?gssapiServiceName=mongodb
Implicit session: session ("id" : UUID("b69312f1-d904-4a8e-a99a-5d1a044023b"))
MongoDB server version: 4.0.10
Welcome to the MongoDB shell.
For interactive help, type "help".
For more comprehensive documentation, see
http://docs.mongodb.org/
Questions? Try the support group
http://groups.google.com/group/mongodb-user
server has startup warnings:
2019-06-23T18:06:29.843-0700 I CONTROL [initandlisten]
2019-06-23T18:06:29.843-0700 I CONTROL [initandlisten] ** WARNING: Access control is not enabled for the database.
2019-06-23T18:06:29.843-0700 I CONTROL [initandlisten] **      Read and write access to data and configuration
2019-06-23T18:06:29.843-0700 I CONTROL [initandlisten] **      will be allowed on this database.
2019-06-23T18:06:29.844-0700 I CONTROL [initandlisten]
To enable MongoDB's free cloud-based monitoring service, which will then receive and display
metrics about your deployment (disk utilization, CPU, operation statistics, etc),
the monitoring data will be available on a MongoDB website with a unique URL accessible to you
and anyone you share the URL with. MongoDB may use this information to make product
improvements and to suggest MongoDB products and deployment options to you.
To enable free monitoring, run the following command: db.enableFreeMonitoring()
To permanently disable this reminder, run the following command: db.disableFreeMonitoring()
```

C. Inserción y consulta de datos

- conexión de Mongo, instancia de Mongo Corriendo dentro de este contenedor docker

```
> db
test
> db.getCollection("test")
test.test
> db.getCollection("test").find("").count()
0
> db.getCollection("test").insert({'name':' virginia aquino'})
```

- Se hace una prueba de Mongo DB, disparar ,para que así la base de datos consiga obtener la prueba de recolección de datos

```
> db.getCollection("test").insert({'name':' virginia aquino'})
2019-06-23T12:26:32.737-0500 E QUERY [js] Error: argument to bsonsize has to be an object :
Bulk/insert@src/mongo/shell/bulk_api.js:611:28
Bulk/this.insert@src/mongo/shell/bulk_api.js:654:20
DBCollection.prototype.insert@src/mongo/shell/collection.js:318:13
@shell:1:11
> db.getCollection("test").insert({'name':' virginia aquino'})
writeResult({ "inserted" : 1 })
```

D. Comparación

E. Documental

La popularidad del término "base de datos orientada a documentos" o "almacén de documentos" ha crecido a la par con el uso del término NoSQL.

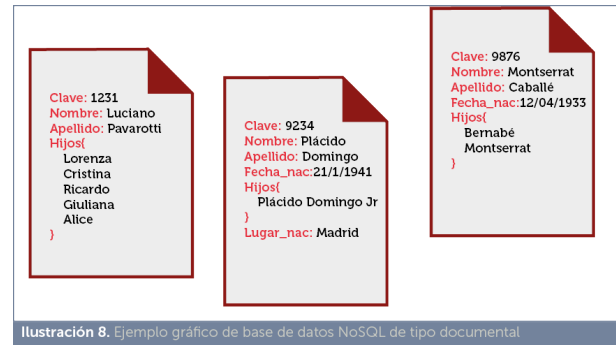
Una base de datos NoSQL documental sirve para gestionar información orientada a documentos o datos semi-estructurados, que se almacenarán en forma de documento.[5]

Normalmente los documentos se estructuran dentro de una serie de contenedores llamadas colecciones (en inglés "collections") que son proporcionados por el sistema gestor documental. Estas utilizan punteros para el acceso de la información almacenada en la base de datos, esto se realiza sin la ayuda de una estructura entidad relación predeterminada, como en el caso de las bases de datos relacionales.

La búsqueda de la información se realiza con base al

contenido del documento.

Las codificaciones más habituales de estos documentos suelen ser XML, YAML o JSON, pero también se pueden almacenar en formato Word o PDF. [4] Las bases de datos de tipo documental son una forma moderna de almacenar datos en formato JSON en lugar de filas y columnas como en las bases de datos relacionales; y esto permite expresar los datos en su forma natural.



Fundamentos del formato JSON :

- Pares de valores clave o atributos : Con el formato JSON se almacena en un par de valores clave. A estos pares se les llama a veces atributos. Las claves son cadenas simples y los valores pueden ser de cualquier tipo.[7]
- Incrustación de objetos: Los valores incluidos en el par de valores clave también pueden ser otros objetos JSON, lo que permite crear una jerarquía de objetos. Colocar objetos JSON dentro de otro objeto JSON se denomina modelo de datos incrustados en base de datos documentales.[7]
- Matrices: El formato JSON trabaja con un lenguaje de programación natural en todos los lenguajes de programación y estructuras de datos que son las matrices, por lo cual se admite almacenamiento de matrices como valores contra una clave. [7]

```

1 [
2   {
3     "year": 2013,
4     "title": "Turn It Down, Or Else!",
5     "info": {
6       "directors": [ "Alice Smith", "Bob Jones" ],
7       "release_date": "2013-01-18T00:00:00Z",
8       "rating": 6.2,
9       "genres": [ "Comedy", "Drama" ],
10      "image_url": "http://ia.media-imdb.com/images/N/05ERNAUT5797A77LUBH094MUP908RL1o5F98ENR7L3KQ7@._v1_SX480_.jpg",
11      "plot": "A rock band plays their music at high volumes, annoying the neighbors.",
12      "actors": [ "David Matthewman", "Jonathan G. Neff" ]
13    }
14  },
15  {
16    "year": 2015,
17    "title": "The Big New Movie",
18    "info": {
19      "plot": "Nothing happens at all.",
20      "rating": 0
21    }
22  }
23 ]
  
```

En la imagen podemos ver un ejemplo de un documento de tipo JSON describe un libro.[2]

Factores que influyen en la aceptación de una BD NoSQL documental :

- **Adopción de un nuevo paradigma:** Este factor de adaptarse al concepto de NoSQL, resulta ser relativamente difícil, ya que la gran mayoría de compañías se encuentran más familiarizadas con las bases de datos relacionales.
- **Curva de aprendizaje:** al ser una tecnología relativamente nueva, esto representa para algunas personas un obstáculo en aprender la utilización de las bases de datos NoSQL
- **Costo-beneficio:** es uno de los factores más importantes, debido al hecho de comparar si los beneficios que se obtendrán son mayores a los costos de implementarlo.

Bases de Datos Documental más conocidas :

- **Amazon DocumentDB** (compatible con MongoDB): Servicio de base de datos de documentos rápido, escalable, de alta disponibilidad y completamente administrado que es compatible con cargas de trabajo de MongoDB.
- **MongoDB** : Base de datos no relacional de código abierto que ofrece soporte para sistemas de almacenamiento de estilo JSON orientados a documentos.
- **Couchbase** : Plataforma de datos de clase empresarial Couchbase incluye Couchbase Server y Couchbase Mobile, y ha sido diseñada para aprovechar toda la potencia de las aplicaciones móviles, IoT y web.
- **RavenDB**: Base de datos orientada a documentos de código abierto escrito en .NET, ofrece un modelo de datos flexible, diseñado para hacer frente a los requerimientos provenientes de los sistemas.

F. Clave-Valor

Las bases de datos clave – valor, llamadas tablas también, se estructuran almacenando la información como si de un diccionario se tratara.

Esta moderna forma de almacenamiento clave – valor se caracterizan por tener una elevada escalabilidad y un rendimiento muy bueno para volúmenes de datos muy grandes, a cambio de ser muy simples y renunciar a funcionalidades que tenemos en otros sistemas como la verificación intrínseca de la integridad de datos, llaves extranjeras y disparadores.

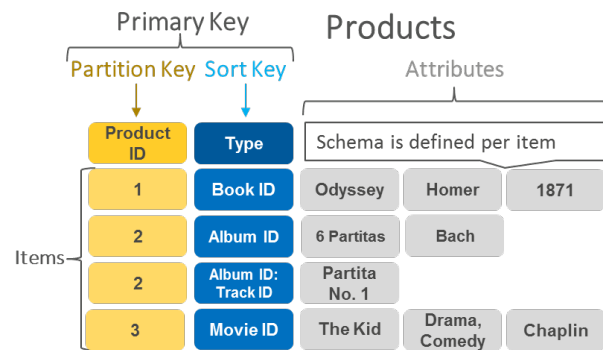
En las BD de clave valor existen los “contenedores”,

y dentro de ellos se pueden almacenar tantos pares clave-valor como se desee.

Las validaciones de los datos se delegan completamente en la aplicación cliente, siendo la base de datos, simplemente el lugar donde se guardan los datos. No se verifican integridades, no se comprueban referencias cruzadas, todo esto se ha de implementar a nivel de aplicación, en el código del cliente.

Características :

- Una base de datos clave-valor almacena datos como un conjunto de pares clave-valor en los que una clave sirve como un identificador único.
- Tanto las claves como los valores pueden ser cualquier cosa, desde objetos simples hasta objetos compuestos complejos.
- Las bases de datos clave-valor son altamente divisibles.
- Permiten el escalado horizontal a escalas que otros tipos de bases de datos no pueden alcanzar.



En la imagen podemos ver un diagrama de datos almacenados como pares clave-valor en DynamoDB.[1]

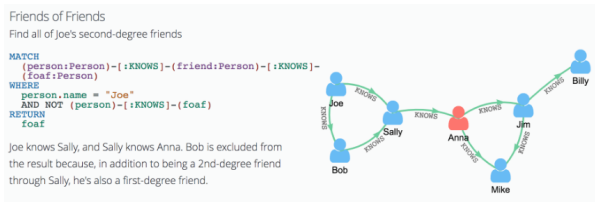
Bases de Datos de Clave-Valor más conocidas :

- **DynamoDB** : El servicio de Amazon permite únicamente crear nuevas tablas en las que alojar los datos, e interactuar con ellas mediante tus aplicaciones a través de APIs.
- **Redis** : Base de datos en memoria de conjuntos clave – valor, que además permite persistir los datos que residen en memoria a disco.
- **Apache Cassandra** : Base de datos no relacional de alto rendimiento utilizada comúnmente.
- **RavenDB**: Base de datos orientada a documentos de código abierto escrito en .NET, ofrece un modelo de datos flexible, diseñado para hacer frente a los requerimientos provenientes de los sistemas.

Los almacenes de llaves valor son útiles para operaciones sencillas, que se basan solo en los atributos clave. Su objetivo es facilitar el acceso a datos.

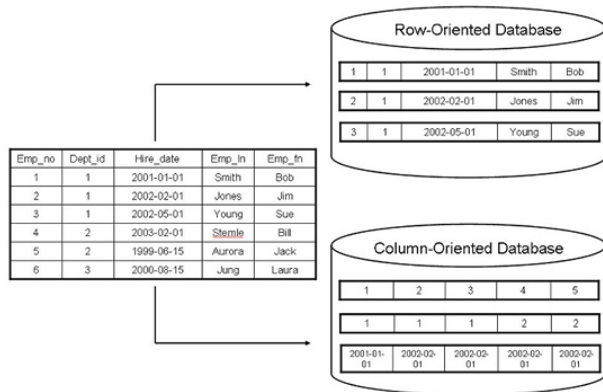
G. Grafos

- Estas bases de datos utilizan el modelo de Grafos
- Se especializan en relaciones
- Las podemos utilizar por ejemplo para guardar puntos de un camino, relaciones de amigos, familia, o cualquier tipo de dato que represente alguna relación
- Entre los motores más conocidos de este tipo se encuentra Neo4j



H. Tabular (Column-Store)

- Este tipo de familia de bases de datos está orientada a grandes cantidades de datos.
- Los datos son almacenados en columnas.
- En una columna tiene múltiples datos.
- Entre los motores más conocidos de este tipo se encuentra Cassandra o HBase.



I. Comparación entre BD Documental y Clave-Valor

	Documentales	Clave-Valor
Formatos	JSON y XML, son los más usados.	String, Numéricos, JSON y estructuras más complejas.
Identificadores	Cada documento consta de un ID único para facilitar métodos de indexación.	Los elementos únicos se identificarán mediante la tupla (bucket, clave).
Ventajas	<ul style="list-style-type: none"> - Simplicidad - Eficiencia - Flexibilidad 	Su organización
Aplicaciones	<ul style="list-style-type: none"> - Datos de sensores provenientes de diferentes fabricantes. - Fichas de clientes con diferentes características. - Catálogos de inventario de productos para una tienda o fábrica. 	<ul style="list-style-type: none"> - Memorias caches de páginas web, donde la URL es la clave y el contenido el valor. - Logs de operaciones, con los timestamp como la clave y el contenido como valor.
Más usadas	<ul style="list-style-type: none"> - MongoDB - CouchDB - CouchBase - MarkLogic 	<ul style="list-style-type: none"> - Riak KV - Redis

V. DISCUSIÓN Y CONCLUSIONES

- NoSQL permite el manejo de grandes volúmenes de datos y la posibilidad de tener un sistema distribuido.
- Las características de las bases de datos NoSQL responden a las necesidades actuales de las diferentes organizaciones, por lo que son una alternativa debido a su capacidad y a la velocidad.
- La integración de ambos, NoSQL y bases de datos relacionales, es un área que debería ser explorada. Las bases de datos NoSQL comercializan funciones de confiabilidad y consistencia para un rendimiento y proceso de escalabilidad extremo, convirtiéndolo en una solución especializada, ya que el número de aplicaciones que pueden depender de las bases de datos NoSQL sigue siendo limitado.

-
- [1] Amazon. La bases de datos clave-valor definida. Accessed: 2019-06-20.
- [2] Amazon. ¿qué es una base de datos de documentos? Accessed: 2019-06-20.
- [3] GeeksforGeeks. Difference between sql and nosql. Accessed: 2019-06-20.
- [4] R. Herranz Gómez. *BASES DE DATOS NOSQL: ARQUITECTURA Y EJEMPLOS DE APLICACIÓN*. 2014.
- [5] S. A. Méndez Aguilar. *USO DE BASES DE DATOS NOSQL DOCUMENTALES PARA CREAR SITIOS WEB DE ALTO RENDIMIENTO*. 2013.
- [6] Nicolas Nisnardi. Base de datos. Accessed: 2019-06-22.
- [7] Kyocera Document Solutions. ¿qué son las bases de datos documentales? Accessed: 2019-06-20.